DOCUMENT RESUME

ED 319 808 TM 015 120

AUTHOR Kino, Mary M.; Roeber, Edward D.
TITLE The State of Assessment Programs: What Are State-Wide Assessment Programs Doing?
PUB DATE Apr 90
NOTE 42p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 17-19, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Educational Assessment; Educational Change; Elementary Secondary Education; Graduation Requirements; Item Banks; Mail Surveys; *Minimum Competency Testing; *National Surveys; *Program Evaluation; Sampling; Scores; *State Programs; *Test Construction
IDENTIFIERS Association of State Assessment Programs

ABSTRACT

A profile of state-wide educational assessment programs is constructed, and the means by which such programs have changed over time and may change in the future are discussed. Data for the study were taken from biennial surveys by the Association of State Assessment Programs, first conducted in 1977. This study covers data on time of survey administration, grade levels, subject areas, type of program, special uses of test results, district controls, sampling procedures, availability of in-house test development capabilities, principal contractors, willingness to share items with other states, item banking, equating and item calibration methods, and assessment of higher order or critical thinking skills. Coding of the data resulted in an interrater reliability rating of 93%. Results indicate that: (1) most programs assess basic skills, use minimum competency programs, or are in some way associated with high school graduation; (2) off-the-shelf tests are often chosen; (3) fewer target grade levels are being tested broadly in many subject matter areas; (4) most programs reflect ongoing changes in educational assessment; and (5) networking among states is crucial. The coding framework used to transcribe data from the survey responses is included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319808

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

MARY M. KINO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The State of Assessment Programs: What Are State-wide Assessment Programs Doing?

Mary M. Kino
Edward D. Roeber

Michigan Department of Education

Running Head: STATE-WIDE ASSESSMENT PROGRAMS

Paper presented at
the National Council on Measurement in Education Annual Meeting
Boston, Massachusetts
April 17, 1990.

TM015120

BEST COPY AVAILABLE

The majority of the Fifty United States conducts state-wide assessment in one form or another. According to the Association of State Assessment Programs (ASAP) biannual surveys, well over 80% of the respondents indicated that their state assessment programs have been and continue to be required by state legislation to conduct state-wide assessment from 1977 to 1989. What have state-wide assessment programs been doing? What areas are being assessed now? In what ways might existing state-wide assessment programs change in the future? Each year, millions of dollars are spent on state-wide assessment programs. To evaluate the level of educational attainment in the state, to provide necessary information to a variety of decision-makers, and to provide educational accountability: These are some of the many reasons most of the Fifty States conduct some form of state-wide assessment.

State-wide assessment began more than twenty years ago to provide accountability to the taxpayers, legislators, educators, and citizens of a state or commonwealth. To confront the concerns and the need to elevate standards of educational achievement, a number of commissions were formed, studies were conducted, and national assessments indexed the status of educational attainment in this country. Assessment sports many faces and has many definitions. In the scope of this study, it is defined, albeit broadly, as the determination of congruence between outcomes and objectives.

Most state departments of education oversee their state's educational activities either directly or indirectly. There is no national curriculum, no centralized department of education which directly monitors the curriculum, instruction, or assessment in each of the Fifty States. Furthermore, few states, if any, have any state curriculum. Many states abdicate curricular and instructional decisions to the local school district or other regional level.

How have state-wide assessment programs changed? Have they, in fact, changed? What current issues in educational assessment have been addressed by these state-wide programs? What are state-wide assessment programs currently doing?

The purposes of this paper are: (1) to construct a profile of state-wide assessment programs; (2) to examine how state-wide assessment programs have changed over time; and (3) to consider the ways in which state-wide assessment programs may change in the future.

Limitations of the Study

Several limitations of this study must be acknowledged. First, the information included in this study depended strictly on available ASAP survey responses. There was no information concerning non-respondent follow-up, and the findings in this paper may provide neither an accurate nor representative picture of state-wide assessment programs when generalized to the Fifty States.

Second, there was no forced coding for missing information. One practice of coding missing information is to infer a response. Using information from other sources, it is sometimes possible to synthesize the most probable response, assuming the respondent unintentionally neglected to provide information. Another practice is to assume that missing information implies the absence of a particular characteristic, in effect assigning a coding value to missing information. For example, no response to the question, "Does your state use basic skills tests," might be interpreted to mean that there are no basic skills tests in use. In these analyses, if there was no response to a question, probable responses were not inferred, no codes were forced, and no values were entered into that field. Only affirmative responses were coded, and it is this set of responses which have been examined.

Third, these surveys are mailed directly to the Chief State School Officers of a state assessment program. Although most survey responses show complete responses to all questions, incomplete responses may be a reflection of the fact that these individuals have insufficient time to complete the survey. It is also possible that the surveys were completed by someone other than a Chief State School Officer, and that the respondent may or may not have had access to the information sought. The reliability and validity of responses, then, is another facet to consider when examining the results.

Fourth, examination of the most recent survey questions indicates that the set of questions has evolved, has been refined, and has a significantly clearer focus when compared to the first survey instrument, which consisted of a request for general information on active state-wide assessment programs. Since the first survey administration in 1977, questions have been phased in and phased out, new ones added, and recurrent ones modified to reflect the current areas of interest. There is, however, a common set of questions which has appeared since fall 1982, and it is this set which provided comparisons over time. In this paper, the variables which were derived from this common set of questions are referenced the "core" variables.

Finally, although response rates have exceeded 60% over the 13-year period, in some years only 20% of the states responded, while more recently, as many as 86% of the 50 states have returned their completed surveys. Rather than report outcomes merely in terms of numbers of states that fall into a particular category (and lacking a better estimator of the true proportion of state-wide assessment programs), percentages of active state-wide programs that fall into a particular category are reported. This will produce an inflated picture when generalizing the results to all Fifty States. It seems, however, more plausible to consider that some portion of the non-respondents' activities and descriptions of assessment programs should also be represented than to assume that a non-response necessarily implies absence of program activities.

The Surveys

The data for this study came from the Association of State Assessment Programs (ASAP) surveys, first conducted in 1977. The specific month of survey administration varied from year to year, but in general, a survey was mailed in the early autumn and again in the late spring of each calendar year. Survey results were summarized several months after data collection and distributed to ASAP survey respondents. All available survey results, from 1977 to the present, have been examined and have been entered in a data base.

The first survey asked respondents to describe their assessment programs. Responses consisted of copies of the state legislation that mandated state-wide assessment, general information fliers designed for the public, or multiple-page essays describing the history of the state assessment program and on-going activities. Since 1977, the survey instrument has been refined and presently consists of a set of systematic and specific questions. Although the actual questions which appear in the survey vary from time to time to reflect interests in emergent educational issues, a set of core questions appears consistently across the surveys. It is this set of core questions from which the variables were derived for coding and for inclusion in the analyses.

Questions

The following questions are representative of the surveys from 1982 to 1989.

(1) Is your state assessment program used to make diagnostic decisions, to assess levels of academic achievement, to set standards of minimal competency, to decide when to promote a student from grade to grade, to endorse high school diplomas, or tied in some other way to high school graduation requirements?

(2) Who comprises the assessment program population?

(3) Approximately how many students are tested at target grade-levels, and what sampling methods, if any, are used to obtain examinee responses?

(4) What subject areas are assessed?

(5) Are tests developed or produced in-house, or are commercially published off-the-shelf tests used?

(6) What new test development projects are being considered, and what technical challenges might they present?

Variables

The collection of all available survey responses was replete with information on external contractors, their responsibilities and contractual services, and the dollar amount of the awarded contract. Also available was information on the annual budget for various state-wide programs and numbers of supervisors, consultants, analysts, and clerical staff. Explanations of the technical problems encountered by the state-wide programs in the

most recent year and plans for the program in the forthcoming year were provided from time to time. More recently, survey responses have come not only from the Fifty States, but foreign provinces and other colonies interested in large-scale assessment. This study included responses only from the Fifty States: All other responses were excluded.

The focuses of this paper are the trends and activities of state-wide programs and the current status of state-wide assessment program activities. The set of variables which are examined are therefore necessarily limited to those for which there are data across a contiguous time. These are referred to as the core variables, and they include the following:

1. time of survey administration;
2. name of responding state;
3. grade levels for which there is state-wide assessment;
4. subject areas tested;
5. type of assessment program;
6. special uses of test results;
7. the availability of district controls;
8. sampling procedures;
9. whether the program has in-house test development activities;
10. names of principal contractors;
11. willingness to share items with other states;
12. item banking;
13. equating and item calibration methods; and
14. assessment of higher order or critical thinking skills.

A coding scheme, interpretation of variables, and special notes have been included in the Appendix.

Analysis

Once the complete set of survey responses was gathered, the responses were coded using a coding scheme. (See Appendix.) An inter-rater consistency index of 47 out of 662 records indicates that among the 662 records, there were inconsistencies in codes for 47 of the records when compared between two independent coders. 93% of the coded records coincided. Discrepancies were resolved, and the recoded information was entered in the data base.

Five data bases were created to facilitate data analysis.

(1) The complete set of codes for the core variables was entered in a data base management program (Microsoft Works, v. 2.00a), and consists of 24 fields per record, 662 records. This master database was sorted to produce two more data bases.

(2) First, the master database was sorted by time of survey administration and then by state. This produced state profiles (a "within-state across time" profile) from 1977 to 1989 for a given state that submitted responses. This state profile provided a way to examine the activities across time for a state-wide assessment program and to detect any changes in the assessment programs for that state. Although for any survey administration from 1977 to 1989, the response rate was always less than 100%, all Fifty States responded to the survey at one time or another.

(3) Second, the database was sorted by state and then by time of survey administration. This produced survey administration profiles (a "between-states within-time" profile) of states that responded during a particular survey administration. The percentage of respondents and any comparative notes were derived from this profile. This data base provided information for trends across time.

(4) Subsequent to the survey administration profile, a year-by-year source of state-wide assessment program funding was formed to examine the proportional allocation of funds over time. Of particular interest was the percentage of states that indicated sole reliance on state-derived funds, federal monies, or a combination of these sources.

(5) Finally, an aggregate time data base was formed. This time data base contains the number of states that fall into any particular coding field for a particular survey administration (e.g., the number of states that used basic skills tests in fall 1982.). Most of the findings regarding trends across time are derived from this aggregate data base.

All figures are based on frequency counts and percentage of incidence among states with active state-wide assessment programs.

Results

Trends across time for a given state: The "within-state across time" profile. The state profiles ("within-state across time") produced three crude categories for level of state-wide program activities: stable, in flux, and growing.

Stable programs are those for which reported target testing grade levels, primary subject areas tested, and type of assessment program and use of tests have remained fairly intact over time. Programs in which specific grade levels may have shifted, say from grade 12 to grade 10, while generally maintaining the extent of testing at other target grade levels, also fell into this category. These states included California, Delaware, Florida, Iowa, Kansas, Michigan, Minnesota, Montana, North Carolina, North Dakota, New Jersey, New Hampshire, New Mexico, Nevada, New York, South Carolina, Utah, Vermont, West Virginia, and Wyoming.

Programs in flux are those that have had their activities interrupted, suspended, or terminated one or more times during the 1977-1989 period. Also falling in this category are those programs for which the target grade levels and/or subject areas and test use have shifted dramatically, say from assessment in grades 1 through 12 to assessment only in grade 8. Approximately 36% of the Fifty States fell into this category and included Alabama, Alaska, Arizona, Connecticut, Georgia, Hawaii, Illinois Louisiana, Maryland, Maine, Missouri, Mississippi, Ohio, Oklahoma, Oregon, Rhode Island, South Dakota, and Wisconsin.

Growing programs are identified as those that show or have shown a steady increase in the number of target grade levels, number of subject areas tested, have begun to engage in in-house test development activities, or have otherwise expanded the scope of their assessment programs. Approximately 24% of the Fifty States fell into this category. These included Arkansas,

Colorado, Idaho, Indiana, Kentucky, Massachusetts, Nebraska, Pennsylvania, Tennessee, Texas, Virginia, and Washington.

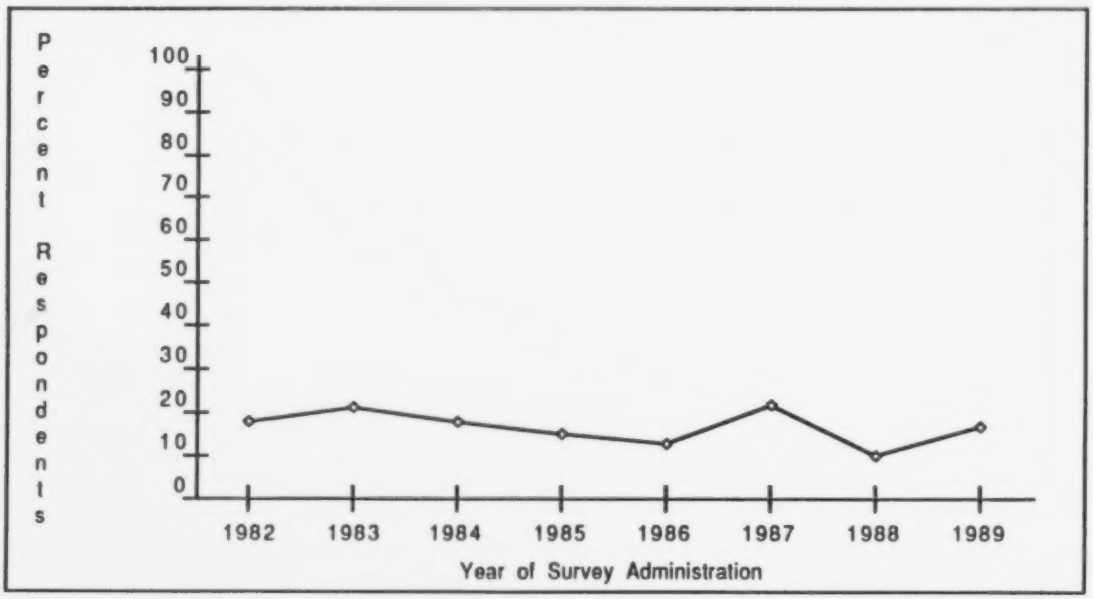
Sources of funding. When examining trends in the data, only data from fall 1982 to fall 1989 were included to form a contiguous time span. Surveys were not regularly administered until fall 1982.

Beginning in fall 1982, annual data on sources of funding for state-wide assessment programs were collected. Some states reported dollar amounts and the sources of each; other states indicated what percentage of their monies came from specific sources; the remaining states either did not supply this information or provided incomplete information. (They may have indicated specific dollar amounts from federal funds, but they did not indicate if this represented all of their funding or if there were additional monies from the state level.)

The results which follow are based only on those responses for which there was complete information. Of the respondents who provided complete information, only 2 of the state assessment programs reported exclusive reliance on federal funding from 1982 through 1989.

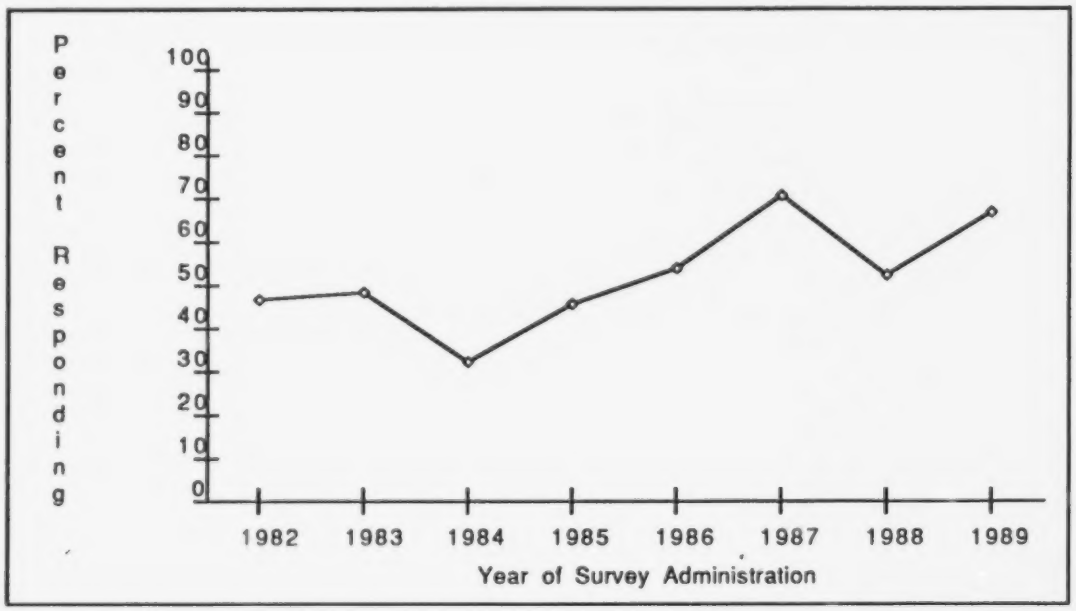
On average, about 17% (8 states) of the complete responses indicated that state program funding came from a combination of state and federal funds, but that the amount of state funds exceeded the amount of federal monies. (See Figure 1.)

FIGURE 1. *Percentage of programs supported by state funds in excess of federal funds.*



On average, from 1982 to 1989 inclusive, 52% (20 states) of respondents indicated their state-wide program funds are derived only from state funds. (Figure 2)

FIGURE 2. *Percentage of programs supported exclusively by state funds.*



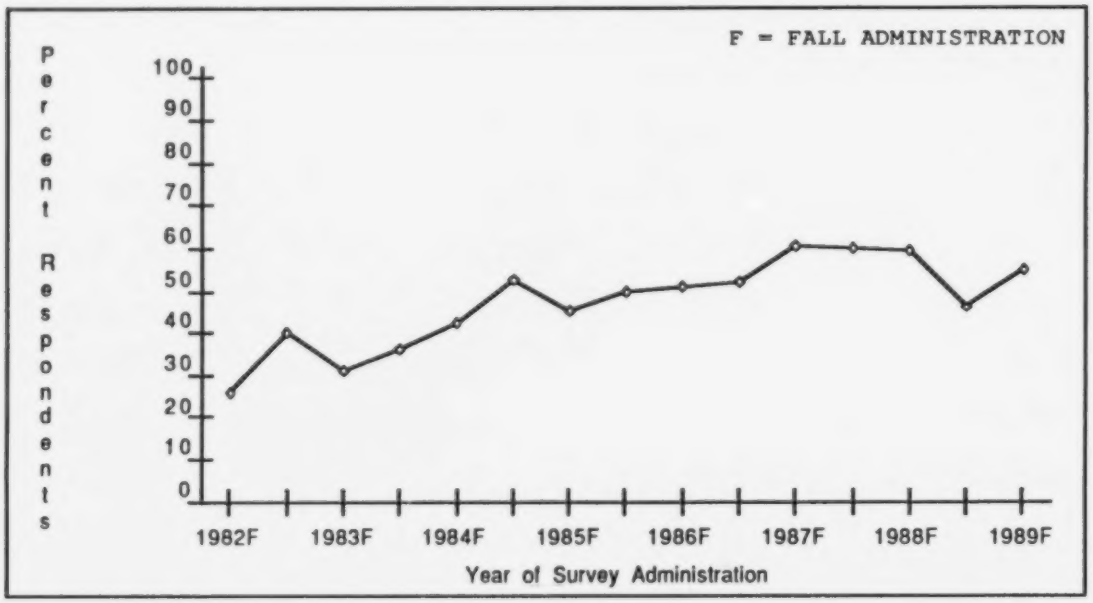
Overall, Figure 2 suggests that state-wide programs are relying more and more on state-derived funds.

Although some respondents (25%) did not indicate the sources of their funds, these findings are based on 75% of the survey responses from states with active state-wide assessment programs. Only one state indicated that in 1984 and in 1985 they received federal support which exceeded state support, and another state indicated that in 1986 state funds were matched by federal monies. The bulk of state assessment programs that provided information on sources of funding relies on state-derived monies.

Profiles of states over time: The "between-states within-time" profile. Most states participate in state-wide assessment in one form or another. Whether to conduct state-wide assessment is the prerogative of a state or commonwealth, and there are states in which there are no state-wide assessment programs. Although the particular forms of assessment programs in these states have changed over the years, they have active large-scale assessment programs either at a district option where the local school district can participate in non-mandatory assessment, choose a test from a state-approved list of tests, determine the particular grade level at which assessment will occur, or report test results to the state level. Most states that do not have mandatory state-wide assessment programs have, either through district option or state mandate, large-scale testing at a local or regional level with the stipulation that test results are to be provided to the state. Such programs have shifted the burden of test administration, scoring, and/or test purchases to the local levels. No more than 11%, on average, of the respondents indicated there is no mandatory state-wide assessment program. There have been fewer than 5 states in which there was no state-wide assessment program during the 1982 to 1989 period.

States in which there are no state-wide assessment programs are not the only ones which may have local school district controls.

FIGURE 3. *Percentage of active programs with district controls.*



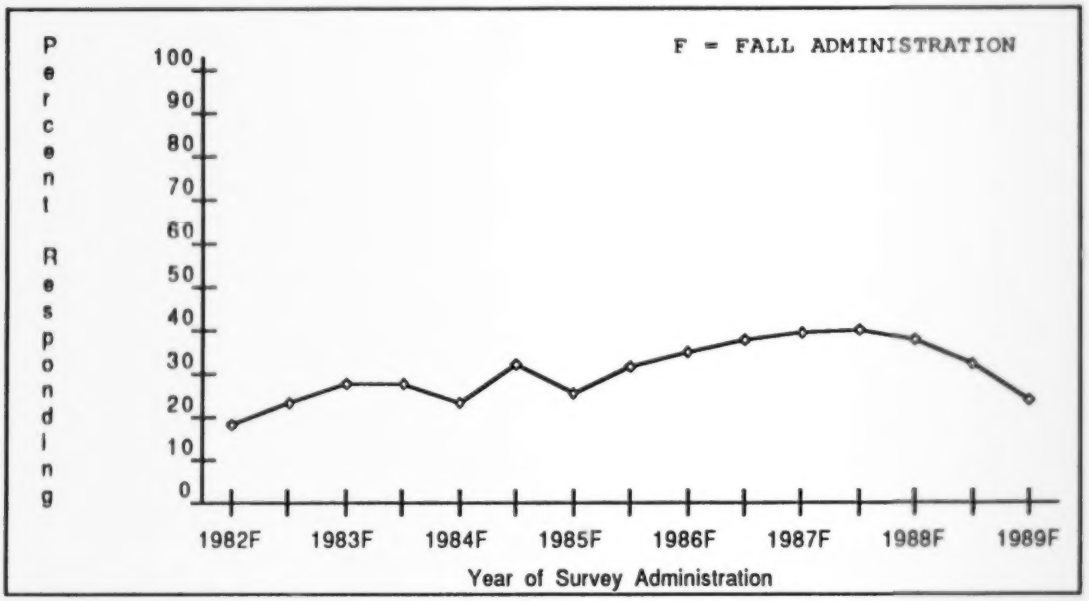
As illustrated in Figure 3, on average, over the 1982 through 1989 period, 47% of the respondents with active state-wide assessment programs indicated they also have provisions for local district controls. School districts can volunteer for optional assessments, select the target grade levels for assessment, or choose from a state-approved list a test or tests to administer. Additionally, these respondents leave the responsibility of determining when to award or deny a high school diploma, when to endorse a high school diploma, or when to promote a student from one grade level to the next.

The National Assessment of Educational Progress (NAEP) conducts its assessment on a cyclic basis using a complex sampling scheme to obtain student responses. Some states have used items from the NAEP in its state-wide assessment program, and others have piggy-backed with the NAEP, have conducted their state-wide assessment concurrent with the NAEP, or have used NAEP in lieu of their state assessment. This is another option that some states with no state-wide assessment program have pursued. On average, over the 1982 to 1989 period, approximately 12% of all respondents have either used NAEP items or have piggy-backed with NAEP assessment.

The scope of state-wide assessment programs differs from program to program. Some conduct state-wide assessment using multiple matrix samples or stratified random samples of students and/or test items to alleviate financial burdens of test development and to reduce testing time per pupil. Other programs conduct assessment on an every-pupil basis.

On average, over the 1982 to 1989 period, 30% of the respondents used sampling methods to obtain pupil responses in their assessment programs. (See Figure 4.)

FIGURE 4. *Percentage of programs using sampling procedures to obtain student responses.*

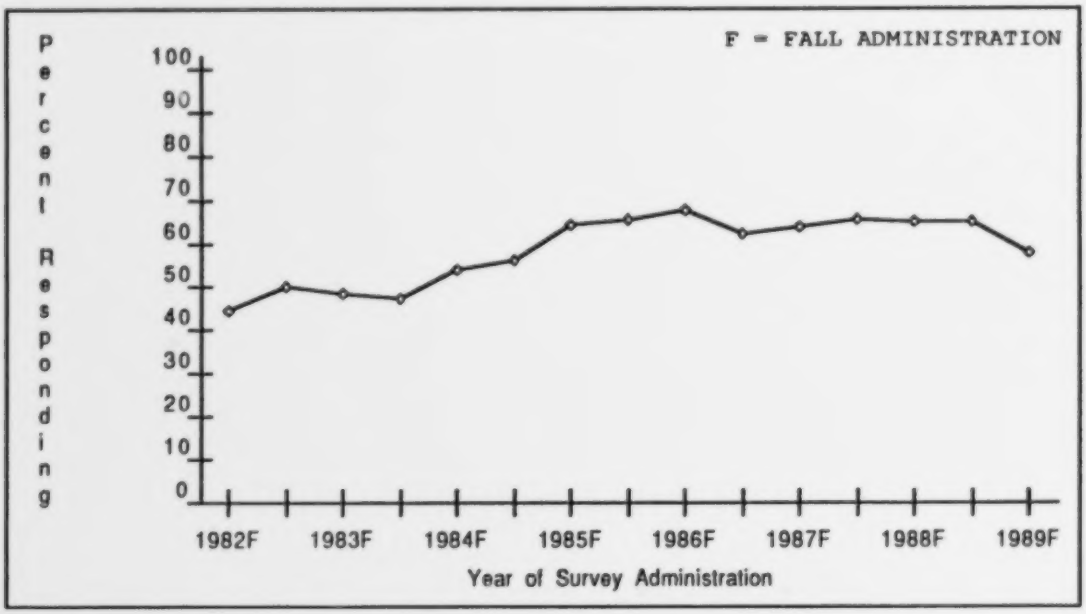


More recently, however, this trend appears to be declining: More state-wide assessment programs appear to be moving toward an every-pupil assessment. Most of these programs have district controls where local districts assume the responsibility of collecting every-pupil data and report the results to the state.

When a program assesses the levels of attainment of objectives, items, or other criteria over a widely varying distribution of ability levels, the differentiation of narrow ranges of abilities becomes increasingly difficult without sacrificing additional classroom time to administer more tests, or

without an increase in financial expenditures to purchase, administer, and score these tests. When asked to describe their assessment programs, as expected, most states indicated they use a basic skills or essential skills test. (Figure 5)

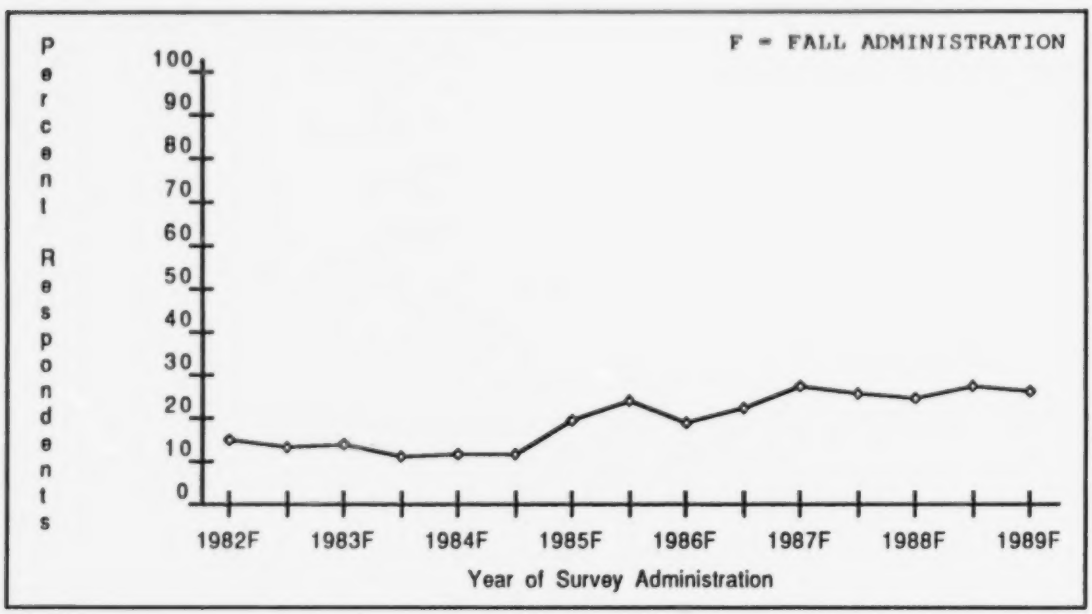
FIGURE 5. *Percentage of programs using basic skills programs.*



On average, over the 1982 to 1989 period, nearly 60% of the respondents indicated they used a basic skills or essential skills test in their state-wide assessment program.

On the average, over the 1982 to 1989 period, 20% of the respondents indicated they used an achievement or mastery skills test in their state-wide assessment program. (Figure 6)

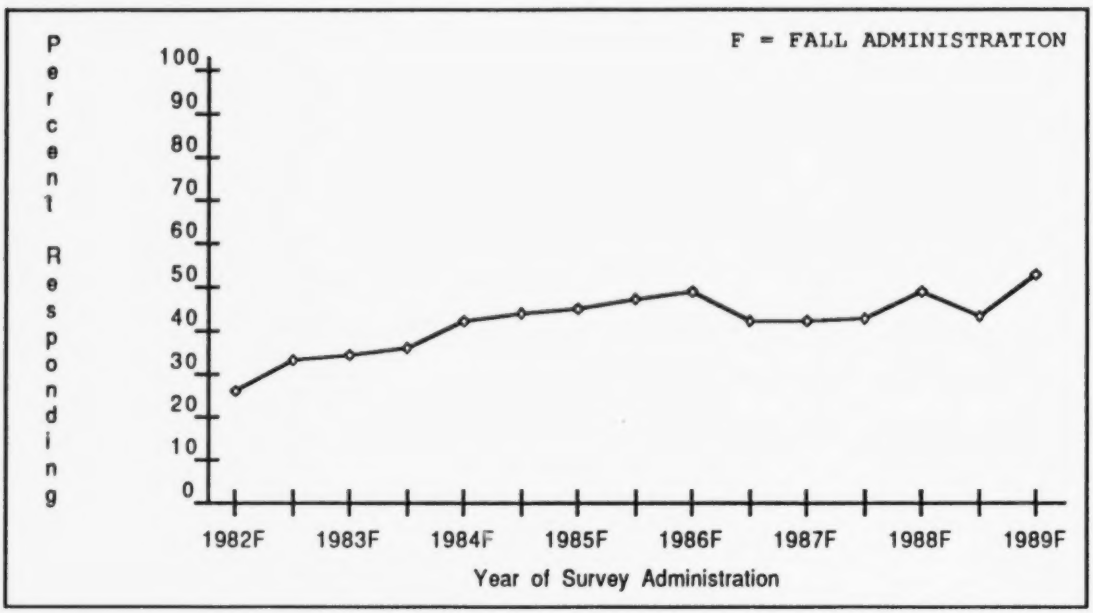
FIGURE 6. *Percentage of programs using achievement or mastery skills programs.*



This percentage is relatively small compared to the percentage of states that use basic skills tests. A subsequent problem when adopting achievement and mastery skills programs is the determination of standards and cut-scores. In many cases, for basic and essential skills, substantive and technical experts often concur on what constitutes basic or essential skills. Agreement in terms of levels of achievement and mastery are typically more difficult to ascertain. The percentage of states that use achievement and mastery skills tests, however, has nearly doubled since 1982, from 15% to nearly 30%.

On average, over the 1982 to 1989 period, more than 40% of the respondents indicated they have a minimum competency program in their state, where the local school districts assume the responsibility of determining whether an individual has attained minimum competency. (See Figure 7.)

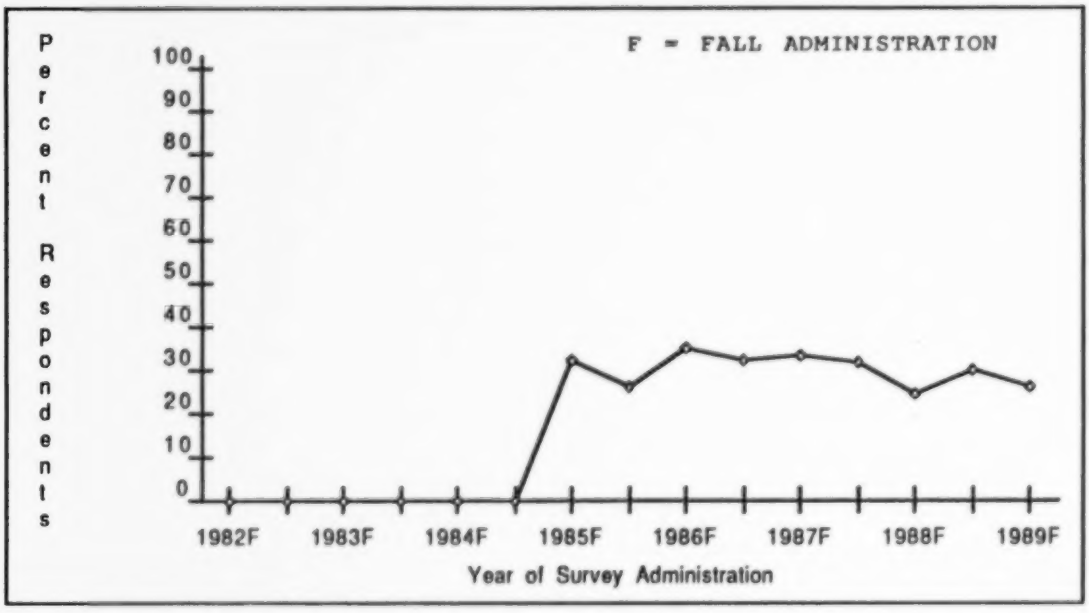
FIGURE 7. *Percentage of programs using minimal competency programs.*



Since fall 1982, the percentage of state programs that use minimum competency assessment programs has increased.

Related to the assessment of achievement levels or mastery skills attainment is the assessment of higher order or critical thinking skills. All of these tend to strive beyond the minimum or the basics.

FIGURE 8. *Percentage of programs that assess higher order thinking skills.*



Over 40% of the state respondents indicated they are either actively assessing the higher order thinking skills in their assessment programs or are currently engaged in activities to promote such assessment. Since 1982, these percentages have also been on the increase. (Figure 8)

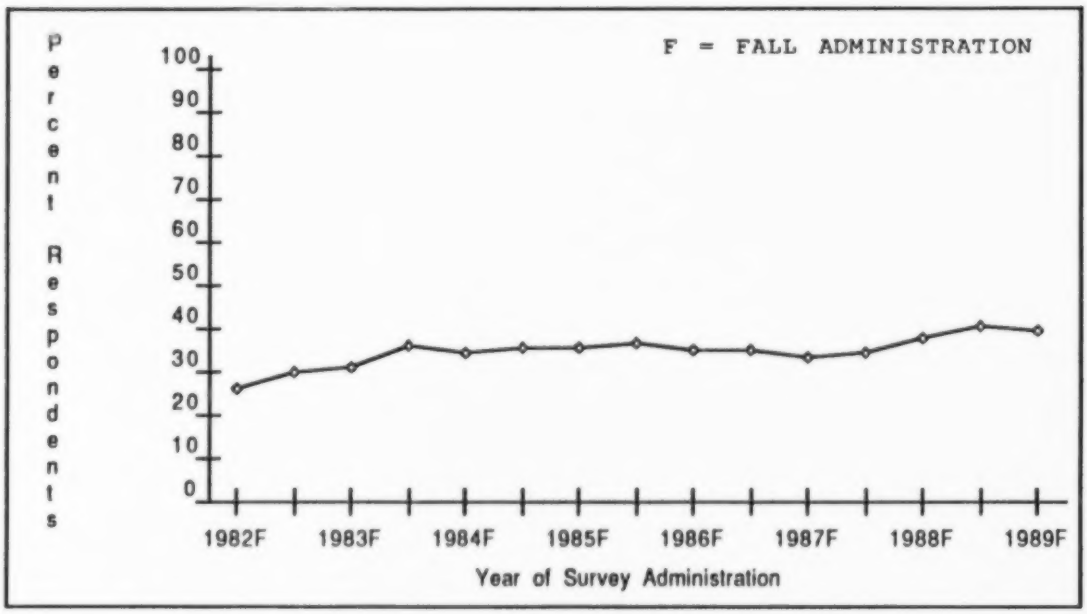
One special use of test results is a kindergarten or elementary first grade screening program or school readiness program. Overall, the percentage of states with active assessment programs that indicated this type of program was less than 15% over the 1982 to 1989 time frame, and no more than 6 states indicated they used this type of assessment program.

Another type of special test use is grade promotion. Less than 10% of the states indicated their assessment programs used test results for decisions regarding promotion from grade to grade. Among the states that indicated they had this type of

program, however, all indicated that local districts had the prerogative to make the grade promotion decisions.

Since 1982, the percentage of states that indicated they implement state-wide assessment programs associated with high school graduation requirements, award of a high school diploma, or high school diploma endorsement has nearly doubled. (See Figure 9.)

FIGURE 9. *Percentage of programs associated with high school graduation requirements.*



On the average, over the period from fall 1982 to fall 1989, nearly 35% of the respondents used test results in conjunction with some aspect of the high school diploma. Like the grade promotional programs, local school districts determine how to enforce this policy.

With the advances in test development technologies, one variable of interest is whether or not states use item banking

technologies. Item banking technologies includes the use of computers (mainframes, pc's, or other stand-alones) or computer software to develop, manage, or catalogue a pool of state-owned items.

FIGURE 10. *Percentage of programs using item banking.*

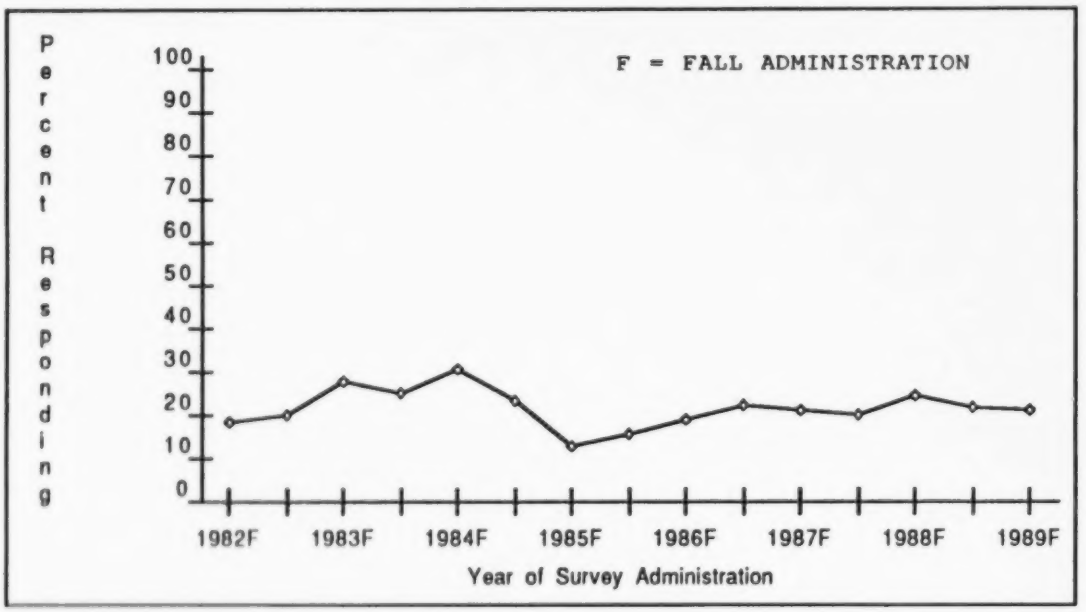
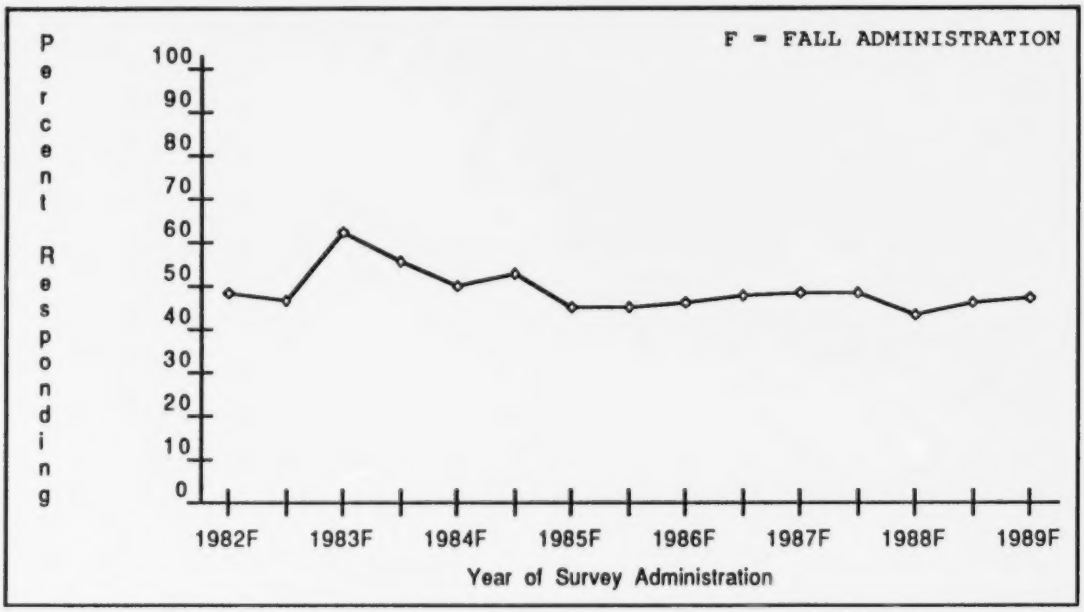


Figure 10 suggests that nearly 22% of the respondents indicated they were using item banking technologies over the 1982 to 1989 period.

Another variable of test development technology is the use of equating methods in state-wide assessment programs. Item calibration and test equating were included in this category. These methods are of particular importance to those state-wide programs that wish to maintain an item pool, to equate old forms

of tests to newer versions, and to perform parallel test construction or engage in some other aspect of in-house test development.

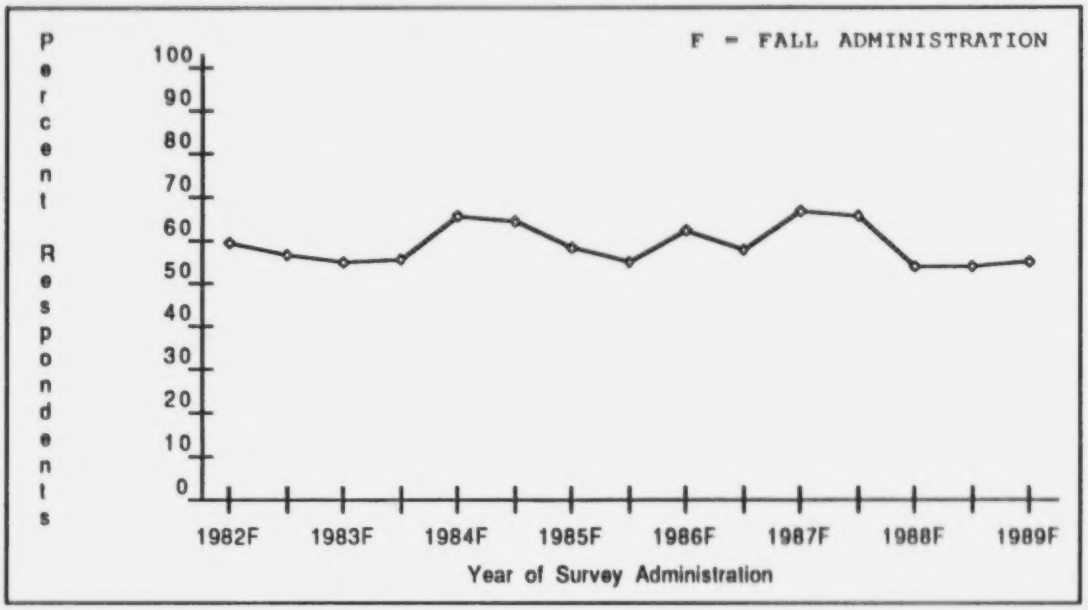
FIGURE 11. *Percentage of programs that use equating methods.*



On the average, over 1982 to 1989, nearly 50% of the respondents indicated they use test equating or item calibration methods. (Figure 11)

To relate the responses of equating and item banking technologies, respondents were asked whether their programs were involved with some aspect of in-house test development, whether through item writing, test development or test production. In-house statistical analyses and other activities with a primary focus on research were excluded from this category.

FIGURE 12. *Percentage of programs involved in in-house test development.*

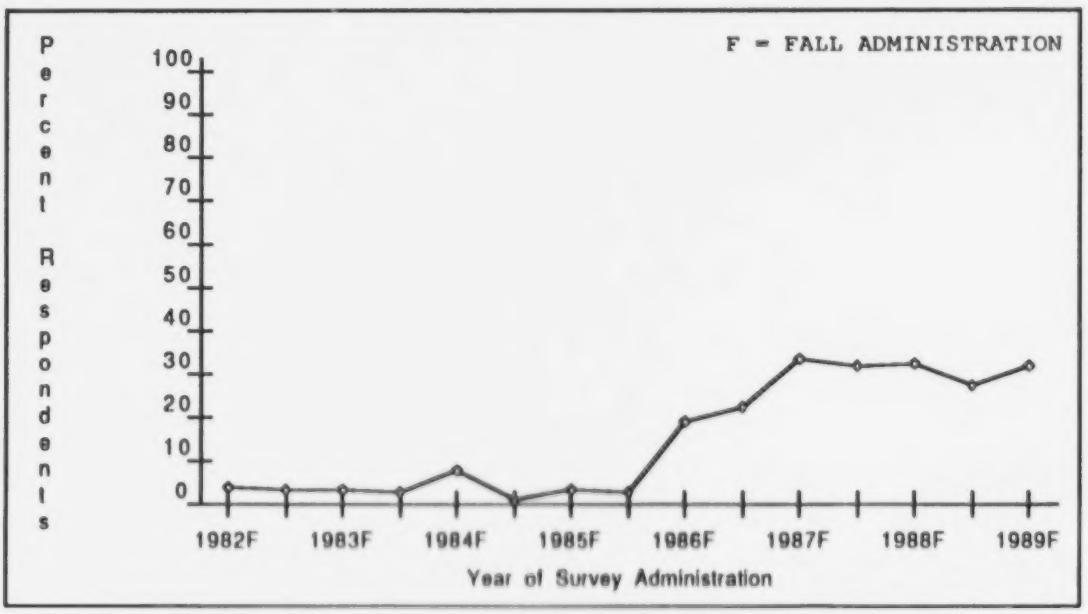


On average, since 1982, nearly 60% of the respondents indicated their programs actively engage in in-house test development activities. (Figure 12)

Respondents were also asked to indicate an ability and willingness to share their items with other state-wide programs. Although a fair percentage of states indicated they were engaged

in in-house test development, Figure 13 shows that most states did not indicate a willingness to share their items with others until 1986.

FIGURE 13. *Percentage of programs willing to share items with others.*



Nearly 30% of respondents have since indicated an interest in sharing items with others.

Finally, a finding which merits greater investigation is the large increase in the use of commercialized off-the-shelf tests in state-wide assessment.

FIGURE 14. *Percentage of programs that use commercially-published off-the-shelf tests.*

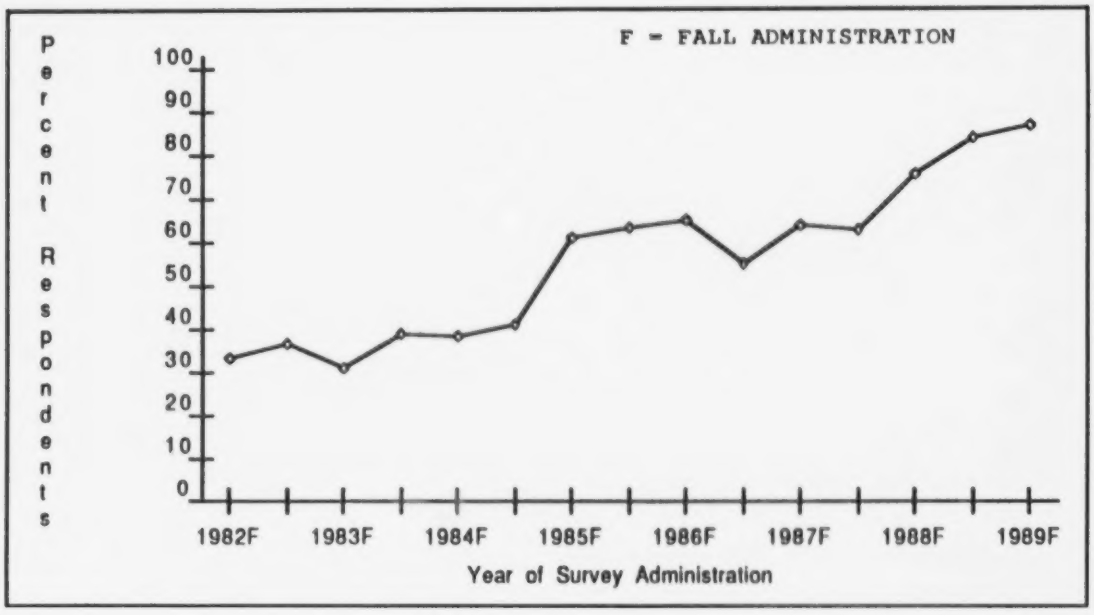


Figure 14 clearly shows a constant increase in the use of commercially published off-the-shelf tests. Nearly 90% of the respondents (33 states out of 38) who have active state-wide assessment programs currently use commercially published tests. A review of which specific commercially published tests are used by which programs indicates that all of the major publishers are represented: There seems to be no great predominance of one commercially published test over another among the states or across time.

Current state of assessment. Table 1, below, summarizes the results from the most recent survey responses for the 1989 calendar year. Figures in parentheses represent the percentage of states that fall into a particular category among those responding. All percentages have been computed based on responses from active state-wide assessment programs except for the percentage of states required to conduct state-wide assessment.

TABLE 1. *Summary of survey responses for calendar year 1989.*

	Spring 1989		Fall 1989	
number of states responding	39		43	
required to conduct state assessment	37	(95%)	38	(88%)
local school district control	17	(46%)	21	(55%)
NAEP-based (items/piggy-back)	6	(16%)	6	(16%)
sampling methods for student responses	12	(32%)	9	(24%)
basic skills	24	(65%)	22	(59%)
achievement or mastery	10	(27%)	10	(26%)
minimum competency programs	16	(43%)	20	(53%)
higher order thinking skills	11	(30%)	10	(26%)
K1 screening/readiness programs	6	(16%)	5	(13%)
grade promotion	5	(14%)	5	(13%)
associated with high school graduation	15	(41%)	15	(39%)
item banking	8	(22%)	8	(21%)
equating	17	(46%)	18	(47%)
in-house test development	20	(54%)	21	(55%)
willing to share items	10	(27%)	12	(32%)
commercially published tests	31	(84%)	33	(87%)

Although more states responded to the fall 1989 survey than to the spring 1989 survey, the percentage of states required by state law to conduct state-wide assessment essentially remains unchanged.

The number of states that indicated that they have provisions for local school district controls has increased from the spring 1989 to fall 1989 administration. More state-wide assessment programs allow local school districts to exercise the option to volunteer for state-wide assessment, to select tests to administer in their districts from a state-approved list, to determine at which grade levels tests will be administered, or to transfer the burden of data collection at the district level.

Relatively few states have indicated they use NAEP-based assessment, use items from the NAEP, or piggy-back with the NAEP for their state-wide assessment. In fact, the same six states have used NAEP-based assessment since spring 1988.

More states appear to be moving toward an every-pupil approach, while at the same time reducing the number of different subject areas in which a student must be tested, or reducing the target grade levels at which testing occurs.

Basic skills assessment programs seems to be the mainstay of state-wide assessment. More states are using minimum competency programs, although not all states use them in conjunction with high school graduation requirements, award of high school diplomas, or endorsement of high school diplomas. Relatively few states are using kindergarten and elementary first grade screening or readiness programs.

Although a fair number of states indicated that they engage in in-house test development activities, few of these have any item banking, do any test equating, or use any item calibration methods, and roughly one dozen of these states are willing to share items with other state-wide assessment programs.

Despite an increasing concern to promote science education in this country, science as a subject area was assessed in only 9 of 37 (barely 25%) of the responding states in spring of 1989, and 10 of the 38 states (slightly more than 25%) in fall of 1989.

Because state-wide assessment programs often provide information to many audiences, other testing programs are often needed to supplement available information from state-wide assessment and to provide additional information. More and more state-wide assessment programs, for example, have begun to use commercially published off-the-shelf testing programs to obtain additional information. Despite the observed increase in use of commercially published off-the-shelf tests, many state-wide assessment programs expressed plans to do their own test development and to move away from the heavy reliance on commercially published off-the-shelf tests.

Summary

According to the most recent ASAP surveys, the majority of the Fifty States are required by state law to conduct state-wide assessment. Most of the state-wide assessment programs assess basic skills, use minimum competency programs, or are in some way associated with high school graduation. A great many of these

state-wide assessment programs use commercially published off-the-shelf tests and provide for local school district controls: It is typically at this level where commercially published tests are used in great predominance. The majority of state assessment programs are engaged in in-house test development activities. Both the increase in use of off-the-shelf tests and in-house test development activities suggest that large-scale testing will continue for some time.

More states are requiring that fewer target grade levels be tested broadly in many subject matter areas, and that subject matter areas are to be assessed in greater depth. States' interests in conducting their own in-house test development suggest that item banking and equating technologies will play an increasingly important role in the future, and that perhaps, there may be an emergence of more innovative technologies in the attainment of that goal.

State-wide assessment programs have been changing and continue to change. Most programs reflect some aspect of the current issues in educational assessment. According to the ASAP surveys, the percentage of states that use commercially published off-the-shelf tests appears to be increasing, and the percentage of states that engage in in-house test development is on the increase. Few of these state-wide assessment programs appear to be diminishing in the intensity or the scope of their testing activities.

As more and more assessment programs begin to develop their own tests, developing a network among other state-wide programs to

share information, to explore innovative methods, and to provide trouble-shooting advice becomes crucial. Not only can networking minimize the duplication of efforts and reduce unnecessary expenditures of time and money to rediscover existing technologies, it can improve the quality of assessment and enhance technological development.

By examining where our state-wide assessment programs have been, where they are now, and where they might be going, we may move toward the first step in disseminating information that we can all use wisely as we plan for the future of educational assessment.

APPENDIX: Coding Sheet

The following is the coding scheme which was used to transcribe information from the survey responses. After all discrepancies in coding between two coders was resolved, the coded information was entered into a data base. Terms in CAPITALS indicate the variable or field name in the data base. A terse definition and interpretation of the terms follow, along with any special notes. Remarks enclosed in brackets [ ] indicate how the variable was coded.

TERM = term of survey administration

The specific month when the survey is mailed differs from year to year. This generally occurs twice annually, once in the fall (September - November) and again in the spring (March - May). This field helps distinguish between the two administrations within a given calendar year.

[Marked either FALL or SPRING.]

YEAR = calendar year of survey administration

The first set of available survey results come from 1977 but are sporadic until the fall 1982 administration. Biannual survey results are available from fall 1982. Any analyses which examine trends in the results use data only from fall 1982 to fall 1989.

[Entered calendar year.]

STATE = state for which respondent is filling out survey

[Coded by two-letter postal abbreviation.]

GRADE LEVELS = target grade levels at which there is any state-wide assessment

It was not possible to ascertain whether all respondents interpreted and identified target grade levels in the state-wide assessment programs in like manner. For example, some state representatives listed all target grade levels for which there was testing, whether the testing occurred at a local school district option with subsequent reporting to the state-wide assessment programs, or whether the assessment occurred at a state-wide level. Records and survey responses suggest, however, that in general, the same individual completed the survey from year to year. If it can be assumed that over repeated administrations, the same individual would interpret the questions with relative consistency, then the primary purpose of coding this information is to examine changes in a given state-wide program across time.

[Listed any and all grade levels identified in the survey.]

SUBJECT = academic subject areas for which there is assessment

Mathematics, reading, language arts, science, social studies, and writing were predominant areas of assessment. Because there was a comparatively lower frequency of testing other areas (e.g., art, music, physical education), these were recoded into an ETC. category in the data base.

[Listed any combination of MTH, RDG, LNG, SCI, SS, WRI, or ETC. as delineated above.]

BASIC SKILLS = those tests which assess skills identified to be basic or essential for functional literacy and functional proficiency

This also includes responses of essential skills as descriptors for assessment programs. Initially, essential skills was a separate descriptor. Because the frequency of identified essential skills programs was relatively low, and since, in some programs, the terms were used interchangeably, essential skills was included in basic skills.

[Indicated either by Y for basic skills/essential skills program or N for no basic skills/essential skills program.]

ACHIEVEMENT = those tests which assess levels of achievement or mastery in a subject area

This also includes responses of mastery skills as descriptors for assessment programs. Like essential skills, mastery skills was a separate descriptor. Some programs used the terms mastery skills and achievement interchangeably. Hence mastery skills is included in achievement.

[Indicated either by Y for achievement/mastery skills program or N for no achievement/mastery skills program.]

MINIMUM COMPETENCY = those tests students must pass to satisfy high school graduation requirements or grade promotion

This descriptor differs from the basic skills descriptor in that there is a particular association, and one which is

potentially high stakes, with the use of these assessment programs.

[Indicated either by Y for minimum competency program or N for no minimum competency program.]

K1 = kindergarten-first grade readiness or screening programs

This descriptor has been used to code programs which test prospective kindergartners and/or elementary first graders for school readiness.

[Indicated either by Y for K1 screening program or N for no K1 screening program.]

GRADE PROMOTION = those programs which use test results in decisions for grade to grade promotions

This descriptor is used only for programs that are used to make decisions for promotion into grades 1 to 12. Kindergarten screening and high school graduation exams are not included in this category.

[Indicated either by Y for grade promotion program or N for no grade promotion program.]

HIGH SCHOOL = those programs which are associated with any aspect of high school graduation

This descriptor has been used to identify those programs which use test results (a) to provide high school diploma endorsement either at the state or local school district levels, or (b) to award high school diplomas.

[Indicated either by Y for association with high school graduation or N for no association with high school graduation.]

NO PROGRAM = no state assessment program

States may have no state-wide assessment program consequent to legislative mandates or lack of financial support to continue assessment activities. These respondents have provided information on their local school district testing programs. Although their responses have been included in the data base, these responses are excluded from analyses of active state-wide programs.

[Indicated either by Y if there is no state-wide assessment program or left blank if there is any state-wide assessment program.]

DISTRICT = programs which provide for local school district controls

These programs allow local school districts to exercise an option to volunteer in their state-wide assessment program; permit districts to choose specific tests and/or target grade levels for assessment, with the stipulation that test results be reported back to the state level; or offer optional assessment at the local school district level.

[Y indicates there is district control, N indicates there is no district control.]

NAEP = programs using items modified from the NAEP, or programs which piggy-back on the NAEP

[Y indicates program uses items modified from the NAEP or uses NAEP as its program, N indicates otherwise.]

SAMPLE = sampling procedures are used to obtain student responses in the state-wide assessment program

State programs in which there is a local school district prerogative to volunteer in state-wide assessment are also considered to use sampling methods. [Y indicates sampling methods are used, N indicates there is census or every-student assessment at the target grade levels.]

NUMBER = number, on average, of students assessed in the state-wide assessment program

IN-HOUSE = programs that engage in some form of their own test development and test production

This descriptor identifies those programs that develop or write their own items, or develop or produce their own test booklets. In-house statistical and item analyses and other aspects of research activities are not included.

[Y indicates the state does its own in-house test development, N indicates otherwise.]

OFF-SHELF = describes those commercially-published, off-the-shelf tests used by a state-wide program

The predominant off-the-shelf tests include the California Achievement Test (CAT), the Comprehensive Test of Basic Skills (CTBS), the Cognitive Aptitude Test (CogAT), the Differential Aptitude Test (DAT), the Iowa Test of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), the Otis-Lennon School Abilities Test (OLSAT), the Stanford Achievement Test (SAT), Science Research Associates series (SRA), and the Test of Academic Progress (TAP).

[Names of off-the-shelf tests were listed as they appeared in the response.]

CONTRACTOR = name of primary contractor for state-wide program

This information was obtained through the profiles of the principal state contractors. Only the names of contractors as supplied in the survey responses were entered into the data base. The primary reason for coding on this field is to examine the term of contract between a contractor and a state.

[Names of principal state contractors were listed as they appeared in the response.]

SHARE = indicates whether a state is willing and able to share test items

[Y indicates a state is willing and able to share items, N indicates otherwise.]

BANKING = item banking technologies

Identifies those programs that use item banking for its state-wide programs.

[Y indicates item banking methods are used, N indicates no item banking at time of survey response.]

HOTS = higher order thinking skills

Included in this field are those programs that use items which assess higher order thinking skills or critical thinking skills. Because the terms have been used interchangeably among many programs, this descriptor addresses both of them.

[Y indicates that the state has items that tap the higher order thinking skills, N indicates otherwise.]

EQUATING = test equating technologies

Identifies those programs which currently do any form of test equating or item calibration. Included in this field are those programs that use item response theoretic (IRT) methods or parallel test construction.

[Y indicates equating, item calibration, IRT, or parallel test construction; N indicates otherwise.]

REPORTS = levels of score reports

Indicates, whenever identified, the levels of score reports provided in the state from state-wide assessment programs. IND indicates score reports at the individual pupil level, CLASS at the classroom level, BLDG at the school building level, DIST at the local school district or other regional level, and STATE at the state level.

[Listed any combination of the above.]